IT & C

ISSN 2821 - 8469, ISSN – L 2821 - 8469, Volumul 1, Numărul 2, Decembrie 2022

Ciclul de viață al inteligenței artificiale

Nicolae Sfetcu

Pentru a cita acest articol: Sfetcu, Nicolae (2022), Ciclul de viață al inteligenței artificiale, *IT & C*, **1**:2, 10-26, DOI: 10.58679/IT46421, https://www.internetmobile.ro/ciclul-de-viata-al-inteligentei-artificiale/

Publicat online: 14.09.2022

ABONARE

Ciclul de viață al inteligenței artificiale

Nicolae Sfetcu

Rezumat

Ciclul de viață al unui sistem al inteligenței artificiale include mai multe faze interdependente, de la proiectarea și dezvoltarea acestuia (inclusiv subfaze precum analiza cerințelor, colectarea datelor, instruire, testare, integrare), instalare, implementare, operare, întreținere și eliminare. Având în vedere complexitatea sistemelor inteligenței artificiale (și în general cele de informații), se pot defini mai multe modele și metodologii pentru a gestiona această complexitate, în special în fazele de proiectare și dezvoltare, cum ar fi dezvoltare de software agilă, cascadă sau spirală, prototipare rapidă și incrementală. Ciclul de viață al inteligenței artificiale definește fazele pe care ar trebui să le urmeze o organizație pentru a profita de tehnicile inteligenței artificiale și în special de modelele de învățare automată pentru a obține valoare practică de afaceri.

Cuvinte cheie: ciclul de viață, inteligența artificială

Abstract

The life cycle of an AI system includes several interrelated phases, from its design and development (including subphases such as requirements analysis, data collection, training, testing, integration), installation, implementation, operation, maintenance and disposal. Given the complexity of artificial intelligence (and information systems in general), several models and methodologies can be defined to manage this complexity, especially in the design and development phases, such as agile, waterfall or spiral software development , rapid and incremental prototyping. The AI lifecycle defines the phases an organization should follow to take advantage of AI techniques and specifically machine learning models to achieve practical business value.

Keywords: life cycle, artificial intelligence

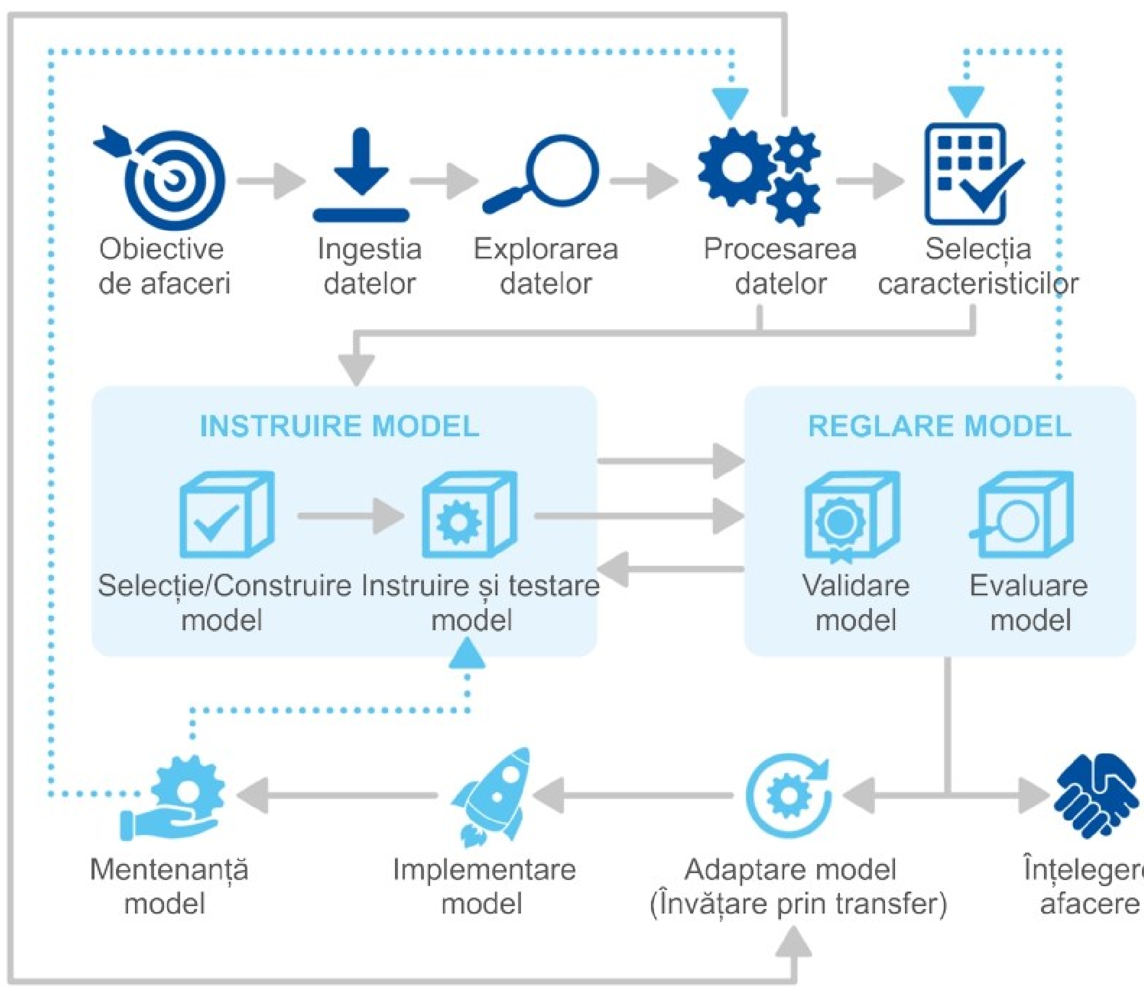
IT & C, Volumul 1, Numărul 2, Decembrie 2022, pp. 10-26
ISSN 2821 - 8469, ISSN – L 2821 - 8469
URL: https://www.internetmobile.ro/ciclul-de-viata-al-inteligentei-artificiale/

Pentru a încadra în mod corespunzător domeniul inteligenței artificiale (AI), este esențial să se urmeze o abordare structurată și metodică pentru a înțelege diferitele sale fațete. Din acest motiv, se poate opta pentru a obține o viziune funcțională a ciclului de viață a sistemelor AI tipice. În consecință, activele implicate (de exemplu, actori, procese, artefacte, hardware etc.), se pot constitui ca bază pentru identificarea amenințărilor (4). Trebuie să se acorde o atenție specială la protecția datelor în contextul AI, o preocupare orizontală în toate etapele ciclului de viață al AI.

Ciclul de viață al unui sistem AI include mai multe faze interdependente, de la proiectarea și dezvoltarea acestuia (inclusiv subfaze precum analiza cerințelor, colectarea datelor, instruire, testare, integrare), instalare, implementare, operare, întreținere și eliminare. Având în vedere complexitatea sistemelor AI (și în general cele de informații), se pot defini mai multe modele și metodologii pentru a gestiona această complexitate, în special în fazele de proiectare și dezvoltare, cum ar fi dezvoltare de software agilă, cascadă sau spirală, prototipare rapidă și incrementală (5). Ciclul de viață AI definește fazele pe care ar trebui să le urmeze o organizație pentru a profita de tehnicile AI și în special de modelele de învățare automată (ML) pentru a obține valoare practică de afaceri. În scopul acestui document, modelele ML sunt utilizate pentru a reprezenta o transformare matematică a datelor de intrare într-un rezultat nou, de ex. utilizați datele de intrare ale imaginii pentru a recunoaște fețele. În schimb, algoritmii sunt utilizați pentru a actualiza parametrii modelului (antrenament) sau pentru a descoperi modele și relații în datele nou furnizate și pentru a deduce rezultatul (6).

Având în vedere gama largă și complexitatea tehnicilor, tehnologiilor, algoritmilor și modelelor implicate în sistemele AI, maparea integrală a acestora într-un singur model de ciclu de viață AI nu este posibilă. Particularitățile sistemelor AI și numeroasele subdomenii ale AI (de exemplu, sisteme de raționament, robotică, AI coecționistă vs simbolică etc.) ar necesita generarea de modele de referință țintite bazate pe tehnologia utilizată. Având în vedere importanța actuală a învățării automate (ML) în utilizarea și implementarea sistemelor AI, am optat pentru a orienta modelul de referință al ciclului de viață AI către ML, pentru a-l face pe de o parte specific și detaliat și, pe de altă parte, abordăm majoritatea sistemelor AI actuale. ML a fost vârful de lance a exploziei AI în ultimii zece ani în ceea ce privește identificarea imaginilor și a vocii.

Ciclul de viață AI

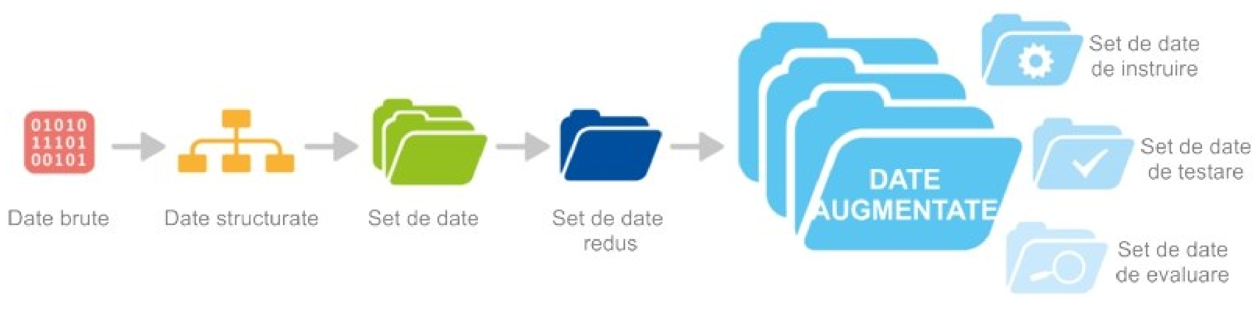


(Model de referință generic pentru ciclul de viață AI)

Pe baza cercetării de birou (7), a fost elaborat un model de referință generic al diferitelor componente găsite în sistemele AI comune, prezentat în figură. Scopul existenței unui model de referință este de a stabili un cadru conceptual care să asigure înțelegerea comună a activelor care compun un sistem AI și relațiile lor semnificative. Acest lucru facilitează alocarea proprietarilor la diferite active pe de o parte și, pe de altă parte, oferă o modalitate sistematică și structurată de analiză a amenințărilor de securitate relevante. Cu condiția ca activele să fi fost definite, amenințările la adresa sistemelor AI pot fi mapate împotriva acestor active și, în urma acestora, pot fi furnizate măsuri de securitate direcționate către proprietarii de active corespunzători.

Datele sunt unul dintre cele mai valoroase active din inteligența artificială; sunt în continuă transformare de-a lungul ciclului de viață AI (8). Figura de mai jos ilustrează transformarea datelor de-a lungul diferitelor etape ale ciclului de viață: Ingestia datelot, Explorarea datelor, Preprocesarea datelor, Evidențierea caracteristicilor, Instruire, Testare și Evaluare. Transformarea datelor de-a lungul ciclului de viață AI implică mai multe alte tipuri de active, cum ar fi actorii implicați, resursele de calcul, software-ul etc., și chiar unele active netangibile, cum ar fi procesele, cultura și modul în care experiența și cunoștințele actorilor pot determina amenințări potențial neintenționale (de exemplu părtinire neintenționată).

Diferitele etape ale ciclului de viață AI se pot descrie punând accent pe diferitele active, procese și actori implicați (9), precum și analizând transformările relevante ale datelor.



(Transformarea datelor de-a lungul etapelor de dezvoltare a ciclului de viață al AI)

Entități implicate în ciclul de viață

Există diferite entități (actori) implicați activ în contextul întregului ciclu de viață al AI. Printre actori se numără designerii AI / designeri de aplicații AI implicați în proiectarea și crearea sistemelor AI. Există, de asemenea, dezvoltatorii AI care dezvoltă și construiesc software-ul și

algoritmii utilizați în sistemele AI, și care lucrează și pentru a le rafina și îmbunătăți. Experiența și capacitatea lor joacă un rol cheie în dezvoltarea sistemelor AI securizate.

Dezvoltatorii și designerii AI lucrează îndeaproape cu specialiștii din știința datelor. Munca acestora ar putea implica asistența la proiectarea și dezvoltarea modelelor AI, sau poate consta în utilizarea unor astfel de modele și analizarea rezultatelor. Mai precis, pecialiștii din știința datelor sunt implicați în colectarea și interpretarea datelor, concentrându-se pe extragerea de cunoștințe și perspective din acele date. Alți actori ai ciclului de viață AI sunt inginerii de date, a căror activitate implică în primul rând extragerea și colectarea datelor din diferite surse, apoi transformarea, curățarea, standardizarea și stocarea acestora. Inginerii de date se concentrează în principal pe proiectarea, gestionarea și optimizarea fluxului de date.

Alți actori importanți ai ciclului de viață AI sunt proprietarii de date (10). Proprietarii de date dețin seturile de date care sunt utilizate fie pentru a instrui/valida sistemele AI, fie pentru a folosi aceste sisteme pentru a îndeplini sarcini. Sunt adesea companii care au propriile seturi de date legate de afacerea lor, la care implementează un sistem AI pentru a îndeplini o sarcină în numele lor. Proprietarii de date pot fi, de asemenea, furnizori / brokeri de date. Aceștia sunt terțe părți care monetizează datele utilizate de sistemele AI, fie în scopuri de instruire, fie pentru a îndeplini diverse sarcini. Aceștia pot include brokerii comerciali de date, care colectează, stochează și vând diferite tipuri de date, în mod legal. Există, de asemenea, rapoarte ale brokerilor de date din zona gri, care adună date despre utilizatori fără ca aceștia să știe că datele lor personale sunt colectate, stocate și vândute (11).

Alți actori ai ciclului de viață AI includ furnizorii de modele, care livrează modele (precum și implementări ale acestora sub formă de biblioteci AI/ML) care au fost deja testate și ajustate. Unii furnizori de modele sunt furnizori de cloud, care oferă modelele ca serviciu, în special utilizarea capabilităților de calcul și analiză a datelor bazate pe AI în cloud. Pe lângă furnizorii de modele, alți actori implică furnizori terți care pot oferi, de asemenea, cadre software și biblioteci terțe, pe care dezvoltatorii le pot folosi pentru instruirea sistemelor AI și hardware specializat de înaltă performanță.

În cele din urmă, există **utilizatorii finali** care folosesc sisteme AI, inclusiv **consumatorii de servicii**. Acestea ar putea fi companii, dintre care multe sunt **utilizatori de modele**. Acestea includ, de asemenea, consumatorii și publicul larg. Utilizatorii finali pot fi și utilizatori ai altor sisteme AI.

Fazele ciclului de viață AI

Definirea scopului

Înainte de a realiza orice dezvoltare de aplicație/sistem AI, este important ca organizația utilizator să înțeleagă pe deplin contextul de afaceri al aplicației/sistemului AI și datele necesare pentru a atinge obiectivele de afaceri ale aplicației AI, precum și valorile de afaceri care vor fi utilizate pentru a evalua gradul în care aceste obiective au fost atinse.

Faza de definire a obiectivelor de afaceri pe scurt: Identificarea scopului comercial al aplicației/sistemului AI. Conectarea scopului cu întrebarea la care trebuie să răspundă modelul AI care va fi utilizat în aplicație/sistem. Identificarea tipului de model pe baza întrebării.

Ingestia datelor

Ingestia datelor este etapa ciclului de viață AI în care datele sunt obținute din surse multiple (datele brute pot fi sub orice formă structurată sau nestructurată) pentru a alcătui puncte de date multidimensionale, numite vectori, pentru utilizare imediată sau pentru stocare pentru a fi accesate și folosite ulterior. Ingestia datelor stă la baza oricărei aplicații AI. Datele pot fi ingerate direct din sursele lor în timp real, într-un mod continuu cunoscut și sub denumirea de streaming, sau prin importul de loturi de date, unde datele sunt importate periodic în macro-loturi mari sau în micro-loturi mici.

Diferite mecanisme de asimilare pot fi active simultan în aceeași aplicație, sincronizând sau decuplând ingerarea în lot și în flux a acelorași fluxuri de date. Componentele de asimilare pot, de asemenea, specifica adnotarea datelor, adică dacă ingerarea este efectuată cu sau fără metadate (dicționar de date sau ontologia/taxonomia tipurilor de date). Adesea, controlul accesului operează în timpul ingerării datelor, modelând starea de confidențialitate a datelor (date personale/non-personale), alegând tehnici adecvate de păstrare a confidențialității și ținând cont de compromisul realizabil între impactul asupra confidențialității și acuratețea analitică. Conformitatea cu cadrul legal aplicabil al UE privind confidențialitatea și protecția datelor trebuie să fie asigurată în toate cazurile.

Statutul de confidențialitate alocat datelor este utilizat pentru a defini Acordul privind nivelul de servicii al aplicației AI (SLA) în conformitate cu cadrul legal aplicabil al UE privind confidențialitatea și protecția datelor, incluzând, printre altele, posibilitatea de inspecție/auditare a autorităților de reglementare competente (cum ar fi Autoritățile de protecție a datelor). Este

important de remarcat că, în ingerarea datelor, poate apărea un conflict de guvernare IT. Pe de o parte, datele sunt compartimentate de către proprietarii săi pentru a asigura controlul accesului și protecția vieții private; pe de altă parte, trebuie să fie integrat pentru a permite analiza. Adesea, pentru articolele din aceeași categorie se aplică politici și reguli diferite. Pentru sursele de date multimedia, protocoalele de acces pot urma chiar și o abordare Digital Right Management (DRM) în care dovada de reținere trebuie mai întâi negociată cu serverele de licență. Este responsabilitatea designerului de aplicații AI să se asigure că ingerarea se face respectând politicile furnizorilor de date privind utilizarea datelor și cadrul legal aplicabil în UE privind confidențialitatea și protecția datelor.

Faza de colectare/ingestie a datelor pe scurt: Identificarea datelor de intrare (dinamice) care trebuie colectate și metadatele de context corespunzătoare. Organizarea asimilării în funcție de cerințele aplicației AI, importând date într-un flux, lot sau multimodal.

Explorarea datelor

Explorarea datelor în inteligența artificială (AI) este etapa în care informațiile încep să fie preluate din datele ingerate. Deși poate fi omisă în unele aplicații AI unde datele sunt bine înțelese, este de obicei o fază a ciclului de viață AI care necesită foarte mult timp. În această etapă, este important să înțelegeți tipul de date care au fost colectate. Trebuie făcută o distincție cheie între diferitele tipuri posibile de date, datele numerice și cele categoriale fiind cele mai proeminente (12), alături de datele multimedia (de exemplu, imagine, audio, video etc.) (13). Datele numerice se pretează la reprezentare grafică și permit calculul statisticilor descriptive și verificarea dacă datele se potrivesc cu distribuțiile parametrice simple precum cea gaussiană. Valorile datelor lipsă pot fi, de asemenea, detectate și gestionate în etapa de explorare. Variabilele categoriale sunt cele care au două sau mai multe categorii, dar fără o ordine intrinsecă. Dacă variabila are o ordonare clară, atunci este considerată ca o variabilă ordinală.

Validarea/explorarea datelor pe scurt: Verificați dacă datele se potrivesc unei distribuții statistice cunoscute, fie prin componentă (distribuții monovariate), fie ca vectori (distribuții multivariate). Estimați parametrii statistici corespunzători.

Preprocesarea datelor

Etapa de pre-procesare a datelor folosește tehnici de curățare, integrare și transformare a datelor. Acest proces are ca scop îmbunătățirea calității datelor care va îmbunătăți performanța și eficiența întregului sistem AI prin economisirea de timp în faza de pregătire a modelelor analitice și prin promovarea unei calități mai bune a rezultatelor. Mai exact, termenul de curățare a datelor desemnează tehnici de corectare a inconsecvențelor, de eliminare a zgomotului și de anonimizare/pseudonimizare a datelor.

Integrarea datelor reunește datele care provin din mai multe surse, în timp ce transformarea datelor pregătește datele pentru a alimenta un model analitic, de obicei prin codificarea lor într-un format numeric. O codificare tipică este o codificare one-hot folosită pentru a reprezenta variabilele categoriale ca vectori binari. Această codificare necesită mai întâi ca valorile categoriale să fie mapate la valori întregi. Apoi, fiecare valoare întreagă este reprezentată ca un vector binar care are toate valorile zero, cu excepția poziției numărului întreg, care este marcat cu 1.

Odată convertite în numere, datele pot fi supuse altor tipuri de transformări: redimensionare, standardizare, normalizare și etichetare (14). La finalul acestui proces, se obține un set de date numerice, care va sta la baza antrenării, testării și evaluării modelului AI.

Deoarece a avea un set de date suficient de mare este unul dintre factorii cheie de succes atunci când se instruiește corect un model, este obișnuit să se aplice diferite tehnici de creștere a datelor acelor seturi de date de antrenament care sunt prea mici. De exemplu, este obișnuit să se includă într-un set de date de antrenament diferite versiuni scalate sau rotite de imagini, care erau deja în acel set de date. Un alt exemplu de tehnică de creștere a datelor care poate fi folosită la procesarea textului este înlocuirea unui cuvânt cu sinonimul său. Chiar și în acele cazuri în care setul de date de antrenament este suficient de mare, tehnicile de creștere a datelor pot îmbunătăți modelul antrenat final. Datele pot fi, de asemenea, augmentate pentru a le crește cantitatea și diversitatea scenariilor acoperite. Augmentarea datelor constă de obicei în aplicarea transformărilor despre care se știe că păstrează etichetele, de exemplu modelul nu ar trebui să-și modifice rezultatul (și anume predicția) atunci când este prezentat cu elementele de date transformate. Augmentarea datelor poate servi la îmbunătățirea performanței unui model și în special a robusteței acestuia la perturbații benigne. O sarcină în care augmentarea datelor este

utilizată în mod implicit este clasificarea imaginilor, unde datele pot fi augmentate, de exemplu, aplicând translații, rotații și filtre de estompare.

Preprocesarea datelor pe scurt: Convertirea datelor ingerate într-un format metric (numeric), integrarea datelor din diferite surse, gestionarea valorilor lipsă/nule prin interpolare, densificarea pentru a reduce dispersitatea datelor, eliminarea zgomotului, filtrarea valorii aberante, modificarea intervalului de reprezentare, anonimizarea/pseudonimizarea datelor, augmentarea datelor.

Selectarea caracteristicilor

Selectarea caracteristicilor (în ingineria generală a caracteristicilor) este etapa în care se reduce numărul de componente sau caracteristici (numite și dimensiuni) care compun fiecare vector de date, prin identificarea componentelor care se consideră a fi cele mai semnificative pentru modelul AI (15). Rezultatul este un set de date redus, deoarece fiecare vector de date are mai puține componente decât înainte (16). Pe lângă reducerea costurilor de calcul, selecția caracteristicilor poate aduce modele mai precise. În plus, modelele construite pe baza datelor de dimensiuni inferioare sunt mai înțelese și explicabile. Această etapă poate fi, de asemenea, încorporată în faza de construire a modelului (de exemplu, la procesarea datelor de imagine sau de vorbire).

Selectarea caracteristicilor pe scurt: Identificarea dimensiunilor setului de date care reprezintă un parametru global, de ex. varianța generală a etichetelor. Datele proiectului sunt stabilite de-a lungul acestor dimensiuni, eliminând pe celelalte.

Selectarea/construirea modelului

Această etapă realizează selecția/construirea celui mai bun model sau algoritm AI (17) pentru analiza datelor. Este o sarcină dificilă, adesea supusă încercărilor și erorilor. Pe baza obiectivului de afaceri și a tipului de date disponibile, pot fi utilizate diferite tipuri de tehnici AI. Cele trei categorii majore identificate în mod obișnuit sunt învățarea supravegheată, învățarea nesupravegheată și modelele de învățare prin întărire. Tehnicile supravegheate tratează datele etichetate: modelul AI este folosit pentru a învăța maparea dintre exemplele de intrare și ieșirile țintă.

Modelele supravegheate pot fi proiectate ca Clasificatori, al căror scop este să prezică o etichetă de clasă, și Regresori, al căror scop este să prezică o funcție de valoare numerică a intrărilor. Aici, câțiva algoritmi obișnuiți sunt Support Vector Machines, Naïve Bayes, Hidden Markov Model, rețele bayesiene și rețele neuronale.

Tehnicile nesupravegheate folosesc date de antrenament neetichetate pentru a descrie și a extrage relații din acestea, cu scopul de a le organiza în clustere, de a evidenția asocierea dintre spațiul de intrare a datelor, de a rezuma distribuția datelor și de a reduce dimensionalitatea datelor. Învățarea prin întărire mapează situații cu acțiuni, prin învățarea comportamentelor care vor maximiza o funcție de recompensă dorită.

În timp ce tipul de date de antrenament, etichetat sau nu, este esențial pentru tipul de tehnică necesar a fi utilizat și selectat, modelele pot fi, de asemenea, construite de la zero (deși acest lucru este destul de puțin probabil), cercetătorul de date proiectând și codificând modelul, cu tehnicile inerente de inginerie software; sau construind un model prin combinarea unei compoziţii de metode (18). Este important de remarcat că selecția modelului (și anume alegerea modelului adaptat la date) poate declanșa o transformare ulterioară a datelor de intrare, deoarece diferite modele AI necesită codificări numerice diferite ale vectorilor de date de intrare.

În general, selectarea unui model include și alegerea strategiei sale de antrenament. În contextul învățării supravegheate, de exemplu, antrenamentul presupune calcularea (o funcție de învățare a) diferenței dintre rezultatul modelului atunci când primește fiecare element de date din set de antrenament D ca intrare și eticheta lui D. Acest rezultat este folosit pentru a modifica modelul pentru a reduce diferența.

Sunt disponibili mulți algoritmi de antrenament pentru minimizarea erorilor, majoritatea bazați pe coborârea gradientului. Algoritmii de antrenament au proprii lor hiperparametri, inclusiv funcția (19) care trebuie utilizată pentru a calcula eroarea modelului (de exemplu, eroarea medie pătrată) și dimensiunea lotului, adică numărul de eșantioane etichetate care urmează să fie alimentate modelului pentru a acumula o valoare a erorii la să fie utilizat pentru adaptarea modelului în sine.

Selecția modelului AI pe scurt: Alegerea tipului de model AI potrivit pentru aplicație. Codificarea vectorilor de intrare a datelor pentru a se potrivi cu formatul de intrare preferat al modelului.

Instruirea modelului

După ce am selectat un model AI, care în contextul acestui model de referință se referă în principal la un model de învățare automată (ML), începe faza de instruire a sistemului AI. În contextul învățării supravegheate, modelul ML selectat trebuie să treacă printr-o fază de antrenament, în care parametrii interni ai modelului, cum ar fi ponderile și părtinirea, sunt învățați din date. Acest lucru permite modelului să înțeleagă datele utilizate și, astfel, să devină mai capabil să le analizeze. Din nou, antrenamentul presupune calcularea (o funcție a) diferenței dintre rezultatul modelului atunci când primește fiecare element de date D al setului de antrenament ca intrare și eticheta lui D. Acest rezultat este folosit pentru a modifica modelul pentru a reduce diferența dintre rezultatul dedus și rezultatul dorit și astfel duce progresiv la rezultate mai precise, așteptate.

Faza de antrenament va alimenta modelul ML cu loturi de vectori de intrare și va folosi funcția de învățare selectată pentru a adapta parametrii interni ai modelului pe baza unei măsuri (de exemplu, pierdere liniară, pătratică, log) a diferenței dintre ieșirea modelului și etichetele. Adesea, setul de date disponibil este împărțit în această etapă într-un set de antrenament, utilizat pentru setarea parametrilor modelului, și un set de testare, în care criteriile de evaluare (de exemplu rata de eroare) sunt înregistrate doar pentru a evalua performanța modelului în afara setului de antrenament. Schemele de validare încrucișată partiționează aleatoriu de mai multe ori un set de date într-un antrenament și o porțiune de testare de dimensiuni fixe (de exemplu, 80% și 20% din datele disponibile) și apoi repetă fazele de instruire și validare pe fiecare partiție.

Instruirea modelului AI pe scurt: Aplicarea algoritmului de antrenament selectat cu parametrii corespunzători pentru a modifica modelul ales în funcție de datele de antrenament. Validarea antrenamentului modelului pe setul de testare conform unei strategii de validare încrucișată.

Ajustarea modelului

Reglajul modelului se suprapune de obicei cu antrenamentul modelului, deoarece reglarea este de obicei luată în considerare în cadrul procesului de antrenament. Am optat pentru separarea celor două etape din ciclul de viață AI pentru a evidenția diferențele în ceea ce privește operarea funcțională, deși cel mai probabil este ca în majoritatea sistemelor AI să fie ambele parte a procesului de instruire.

Anumiți parametri definesc concepte de nivel înalt despre model, cum ar fi funcția sau modalitatea lor de învățare, și nu pot fi învățați din datele de intrare. Acești parametri speciali, adesea numiți hiper-parametri, trebuie configurați manual, deși în anumite circumstanțe pot fi reglați automat prin căutarea în spațiul parametrilor modelului (20). Această căutare, numită optimizare cu hiper-parametri (21), este adesea efectuată folosind tehnici clasice de optimizare, cum ar fi căutarea în grilă, dar pot fi utilizate căutarea aleatorie și optimizarea bayesiană. Este important de remarcat faptul că etapa Ajustarea modelului folosește un set de date special (numit adesea set de validare), distinct de seturile de antrenament și test utilizate în etapele anterioare. De asemenea, poate fi luată în considerare o fază de evaluare pentru a estima limitele rezultatelor și pentru a evalua modul în care modelul s-ar comporta în condiții extreme, de exemplu, prin utilizarea seturilor de date greșite/nesigure. Este important de remarcat faptul că, în funcție de numărul de hiper-parametri care trebuie ajustați, încercarea tuturor combinațiilor posibile poate să nu fie fezabilă.

Ajustarea modelului AI pe scurt: Aplicarea adaptării modelului la hiper-parametrii modelului AI antrenat folosind un set de date de validare, în funcție de condiția de implementare.

Învățarea prin transfer

În această fază, organizația utilizatorului se aprovizionează cu un model AI pre-antrenat și pre-ajustat și îl folosește ca punct de plecare pentru formarea ulterioară pentru a obține o convergență mai rapidă și mai bună. Acesta este de obicei cazul când sunt disponibile puține date pentru antrenament. Trebuie remarcat faptul că toți pașii descriși mai sus (reglare, testare etc.) se aplică și pentru învățarea prin transfer. Mai mult, deoarece în multe cazuri învățarea prin transfer este aplicată, se poate considera învățarea prin transfer ca parte a fazei de antrenament a modelului, având în vedere că învățarea prin transfer servește de obicei ca punct de plecare al algoritmului de antrenament. Pentru a asigura un domeniu mai larg, distingem învățarea prin transfer într-o fază distinctă din ciclul de viață AI prezentat aici.

Învățarea prin transfer pe scurt: Obținerea unui model AI pregătit în prealabil în același domeniu de aplicație și aplicarea instruirii suplimentare, după cum este necesar, pentru îmbunătățirea acurateții în producție.

Implementarea modelului

Un model de învățare automată va aduce cunoaștere unei organizații numai atunci când predicțiile sale devin disponibile pentru utilizatorii finali. Implementarea este procesul de preluare a unui model instruit și de punere la dispoziție utilizatorilor.

Implementarea modelului pe scurt: Generarea unei încadrări în producție a modelului ca software, firmware sau hardware. Implementarea încadrării modelului în edge sau cloud, conectând fluxurile de date din producție.

Întreținerea modelului

După implementare, modelele AI trebuie monitorizate și menținute în mod continuu pentru a gestiona schimbările de concept și potențialele devieri de concept care pot apărea în timpul funcționării lor. O schimbare de concept are loc atunci când semnificația unei intrări în model (sau a unei etichete de ieșire) se schimbă, de ex. din cauza reglementărilor modificate. O deviere a conceptului apare atunci când schimbarea nu este drastică, ci apare încet. Deriva se datorează adesea încrustării senzorului, adică evoluției lente în timp a rezoluției senzorului (cea mai mică diferență detectabilă între două valori) sau a intervalului general de reprezentare. O strategie populară pentru a gestiona întreținerea modelului este reînvățarea bazată pe ferestre, care se bazează pe puncte de date recente pentru a construi un model ML. O altă tehnică utilă pentru întreținerea modelului AI este testarea înapoi. În cele mai multe cazuri, organizația utilizatoruluir știe ce s-a întâmplat după adoptarea modelului AI și poate compara predicția modelului cu realitatea. Acest lucru evidențiază schimbările de concept: dacă un concept de bază se schimbă, organizațiile văd o scădere a performanței. O altă modalitate de a detecta aceste devieri de concept poate determina caracterizarea statistică a setului de date de intrare utilizat pentru antrenamentul modelului AI, astfel încât să fie posibil să se compare acest set de date de antrenament cu datele de intrare curente în ceea ce privește proprietățile statistice. Diferențele semnificative între seturile de date pot indica prezența unor potențiale deviații de concept care pot necesita efectuarea unui proces de reînvățare, chiar înainte ca rezultatul sistemului să fie afectat semnificativ. În acest fel, procesele de recalificare/reînvățare, care pot fi consumatoare de timp și resurse, pot fi efectuate numai atunci când este necesar și nu periodic, ca în strategiile de reînvățare bazate pe ferestre menționate mai sus. Întreținerea modelului reflectă, de asemenea, nevoia de a monitoriza

obiectivele și activele de afaceri care ar putea evolua în timp și, în consecință, să influențeze modelul în sine.

Întreținerea modelului pe scurt: Monitorizarea rezultatelor inferenței ML ale modelului AI implementat, precum și datele de intrare primite de model, pentru a detecta posibile modificări sau derive ale conceptului. Reantrenarea modelului atunci când este necesar.

Înțelegerea afacerii

Construirea unui model AI este adesea costisitoare și întotdeauna necesită timp. Prezintă mai multe riscuri de afaceri, inclusiv nerespectarea unui impact semnificativ asupra organizației utilizatorului, precum și lipsa termenelor limită în producție după finalizare. Înțelegerea afacerii este etapa în care companiile care implementează modele AI obțin o perspectivă asupra impactului AI asupra afacerii lor și încearcă să maximizeze probabilitatea de succes.

Înțelegerea afacerii pe scurt: Evaluarea propunerilor de valoare a modelului AI implementat. Estimarea (înainte de implementare) și verificarea (după implementare) a impactului său asupra afacerii.

Note

1. A se vedea https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines, aprilie 2019
2. Evident, modelele de cutie albă sunt, de asemenea, susceptibile la atacuri cibernetice, deoarece adversarii au informații la scară largă pentru a adapta atacurile.
3. Aceasta se referă atât la atacurile fizice asupra sistemelor AI, cât și la robustețea sistemelor AI împotriva variațiilor și evenimentelor care apar în mod natural.
4. Aici considerăm că sursele de date pentru AI au fost protejate și sunt considerate a fi sigure. În abordarea noastră, ciclul de viață al aplicației AI este considerat un model generic pentru fundamentul identificării activelor și amenințărilor, și nu este conceput ca o declarație. Buclele de feedback prezentate nu sunt exhaustive, deoarece cazuri de utilizare diferite pot urma drumuri diferite și omite unele dintre fazele ciclului de viață generic. Hărțile mentale au fost incluse ca un prim pas către un model de referință complet.
5. A se vedea https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment
6. Peisajul amenințărilor presupune înțelegerea de bază a terminologiei și conceptelor AI.
7. Inclusiv lucrările deja menționate de la EC JRC, EC AI HLEG, EDA, ETSI ISG SAI, NIST, Stiftung Neue Verantwortung, Mcrosoft (https://docs.microsoft.com/en-us/security/engineering/threat-modeling-aiml), Berryville Institute of Machine Learning (https://berryvilleiml.com/) și BSI (https://doi.org/10.3389/fdata.2020.00023).
8. În ceea ce privește categoriile de date și proveniența datelor, distingem între următoarele.

   a. Date auto-raportate, furnizate voluntar de un operator „de încredere” (de exemplu, AIS pentru o navă sau ADS-B pentru o aeronavă, date cooperative și guvernamentale).
   b. Datele observate colectate de sisteme „securizate” active sau pasive (de exemplu IDS, senzori, RFID, camere, IoT în general, radare), integritatea datelor depinde de o varietate de parametri (rezoluție, interval, reîmprospătare, latență, condiții de mediu, dimensiune, orientare, caracteristici electromagnetice).
   c. Registre de informații și baze de date: conțin informații care leagă date (ID-uri aeronave sau nave, ID-uri umane din sistemele moștenite civile, ID-uri obiecte inteligente din industrii) cu detalii despre structura acesteora, construcție, aspect, istoric și interacțiuni, activitate, social media din sursele de internet libere și deschise (ex. Twitter, Youtube, Facebook, WhatsApp, Media, Open DB) sunt de asemenea incluse în această categorie.
9. Modelul de referință detaliază fazele tipice, diferite ale ciclului de viață AI. O referire demnă de remarcat trebuie făcută la soluțiile automate de învățare automată (oferite de mai mulți furnizori) care cuprind marea majoritate a etapelor ciclului de viață AI pentru a facilita dezvoltatorii de produse. În ciuda numeroaselor inițiative de cercetare și comerciale pentru dezvoltarea unor mecanisme și instrumente automate eficiente de învățare automată, au fost identificate multe provocări, inclusiv probleme de transparență (funcționare în cutie neagră), reproductibilitate limitată etc.
10. De reținut că, în cazul datelor cu caracter personal, rolul proprietarilor de date este echivalent cu cel al operatorilor de date.
11. Evident, dacă apar astfel de cazuri, atunci există o lipsă clară de conformitate cu prevederile GDPR și o analiză juridică suplimentară (în afara domeniului de aplicare al acestei lucrări) este foarte recomandată.
12. Discuția se referă în principal la date nunerice, tabelare. Cu toate acestea, trebuie menționat că sistemele AI pot folosi și alte tipuri de date, de ex. discurs, imagini. Acestea sunt, de asemenea, numerice, dar verificările de corectitudine au un grad avansat de complexitate, pentru care nu se efectuează nicio explorare a datelor așa cum este descrisă aici.
13. Datele multimedia sunt date complexe care sunt foarte relevante în contextul învățării profunde.
14. Re-scalarea este utilizată pentru a se asigura că toate variabilele sunt exprimate pe aceeași scară, deoarece unele metode pot trece cu vederea variabilele cu intensitate mai mică. Standardizarea este utilizată pentru a modifica media unei distribuții de valori la 0, în timp ce normalizarea mapează datele la un interval de reprezentare compact (de exemplu, intervalul (0, 1), împărțind toate valorile la maxim). Etichetarea (realizată de experți umani sau de alte aplicații AI) asociază fiecare element de date la o categorie sau o predicție.
15. Modelele de învățare automată sunt algoritmi antrenați cu date istorice care descoperă modele și relații și construiesc modele matematice folosind aceste descoperiri.
16. Este de remarcat faptul că nu este întotdeauna cazul. În special, în abordările recente de învățare profundă care iau în considerare abordări end-to-end de învățare profundă, în care nu se realizează nicio procesare a caracteristicilor.
17. Stuart J. Russell și Peter Norvig, „*Artificial Intelligence: A Modern Approach*”, Prentice Hall Press. ISBN:978-0-13-604259-4

18. Prin combinare de metode ne referim la ansamblu de modele care constă în combinarea rezultatelor mai multor modele pentru a profita de avantajele diferitelor abordări, cu prețul unei complexități mai mari.
19. În învățarea profundă, unde sunt concepute funcții de pierdere posibil extrem de complexe și sunt un element cheie al procesului de formare.
20. Reglarea hiper-paratametrelor este adesea o sarcină dificilă, având în vedere că spațiul hiper-parametrilor este de obicei imens, iar procesul necesită o cantitate mare de timp și resurse de calcul. Mai mult, trebuie remarcat faptul că acest tip de reglare necesită o reinstruire frecventă a modelului.
21. Trebuie remarcat faptul că acest proces este foarte costisitor din punct de vedere computațional și tinde să fie limitat, în special în aplicațiile de învățare profundă, unde antrenamentul poate dura zile sau săptămâni.

Referințe

L.A. Adamic – N. Glance, The political blogosphere and the 2004 US election: divided they blog, Proceedings of the 3rd International Workshop on Link discovery 2005, pp. 36-43

C. Aslay et al., Maximising the diversity of exposure in a social network, IEEE International Conference on Data Mining 2018, pp. 863-868

A. Caliskan Islam et al., Semantics derived automatically from language corpora necessarily contain human biases, arXiv preprint arXiv:1608.07187 2016

R. Carter – H.V. Auken., Small firm bankruptcy, Journal of Small Business Management 2006, 44: pp. 493-512

L. De Biase, Homo pluralis. Essere umani nell'era tecnologica. Codice, Torino 2016

K. Garimella et al., Reducing controversy by connecting opposing views, Proceedings of the Tenth ACM International Conference on Web Search and Data Mining 2017, pp. 81-90

R. Guidotti et al., A survey of methods for explaining black box models, ACM computing surveys (CSUR) 2018, pp. 1-42

R. Hegselmann – U. Krause, Opinion dynamics and bounded confidence: models, analysis and simulation, «Journal of Artificial Societies and Social Simulation» V (2002) 3

S. Lowry – G. Macpherson, A blot on the profession, British medical journal (Clinical research ed.) 1988, pp. 657

F. Pasquale, The black box society, Harvard University Press 2015

D. Pedreschi et al., Discrimination-aware data mining, Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining 2008, pp. 560-568

D. Pedreschi et al., Meaningful explanations of Black Box AI decision systems. Proceedings of the 33rd AAAI Conference on Artificial Intelligence 2019, 9780-9784

S. Plous, The Psychology of Judgment and Decision Making. McGraw-Hill, New York 1993

M.T. Ribeiro et al., "Why should I trust you?" Explaining the predictions of any classifier, Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining 2016, pp. 1135-1144

A.L. Schmidt et al., Polarisation of the vaccination debate on Facebook, National Center for Biotechnology Information 36 (2018) 25, pp. 3606-3612

A. Sirbu et al., Algorithmic bias amplifies opinion fragmentation and polarisation: A bounded confidence model, «PLoS ONE» 14(3), 2019

J. Surowiecki, The wisdom of crowds, Anchor Books, New York 2004

Artificial Intelligence for Europe, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and social committee and the committee of the Regions, Brussels 2018

Divina Frau-Meigs, Societal costs of “fake news” in the Digital Single Market, Study requested by the IMCO committee, 2018

Žiga TURK, Technology as Enabler of Fake News and a Potential Tool to Combat It, In-Depth Analysis requested by the IMCO committee, 2018

XAI (2019-2024, ERC Advanced Grants 2018) Science and technology for the explanation of AI decision making. https://xai-project.eu/

SoBigData (2015-2024, H2020-Excellent Science Research Infrastructures) Integrated Infrastructure for Social Mining & Big Data Analytics. A research infrastructure at the second stage of “Advanced community”, aggregating 32 partners of 12 EU Countries. http://www.sobigdata.eu/

Humane-AI (2019-2020, H2020-FETFLAG-2018-01 Coordination Action) Toward AI Systems That Augment and Empower Humans by Understanding Us, our Society and the World Around Us. https://www.humane-ai.eu/

Sursa: Sfetcu, Nicolae (2020). *Analitica rețelelor sociale*, MultiMedia Publishing, ISBN 978-606-033-704-1, https://www.telework.ro/ro/e-books/analitica-retelelor-sociale/.